

Problemy przechowywania danych cyfrowych w bibliotekach

Dorota Witczak, Krzysztof Sobkowiak

Biblioteka Główna Uniwersytetu Pedagogicznego im. KEN w Krakowie

Traditionally, preserving things meant keeping them unchanged; however, our digital environment has fundamentally changed our concept of preservation requirements. If we hold on to digital information without modifications, accessing the information will become increasingly more difficult, if not impossible.[1]

Streszczenie

W artykule przedstawiono problemy z przechowywaniem danych cyfrowych w bibliotekach, które są nie tylko nabywcami danych, ale i ich wytwórcami. Przybliżono wybrane światowe i polskie projekty bezpiecznej i długoterminowej archiwizacji. Zwrócono uwagę na konieczność wypracowania ogólnopolskiej strategii i procedur właściwego przechowywania i zarządzania cyfrowymi zasobami, aby zapewnić im długoterminową użyteczność oraz znacząco obniżyć koszty wieczystej archiwizacji.

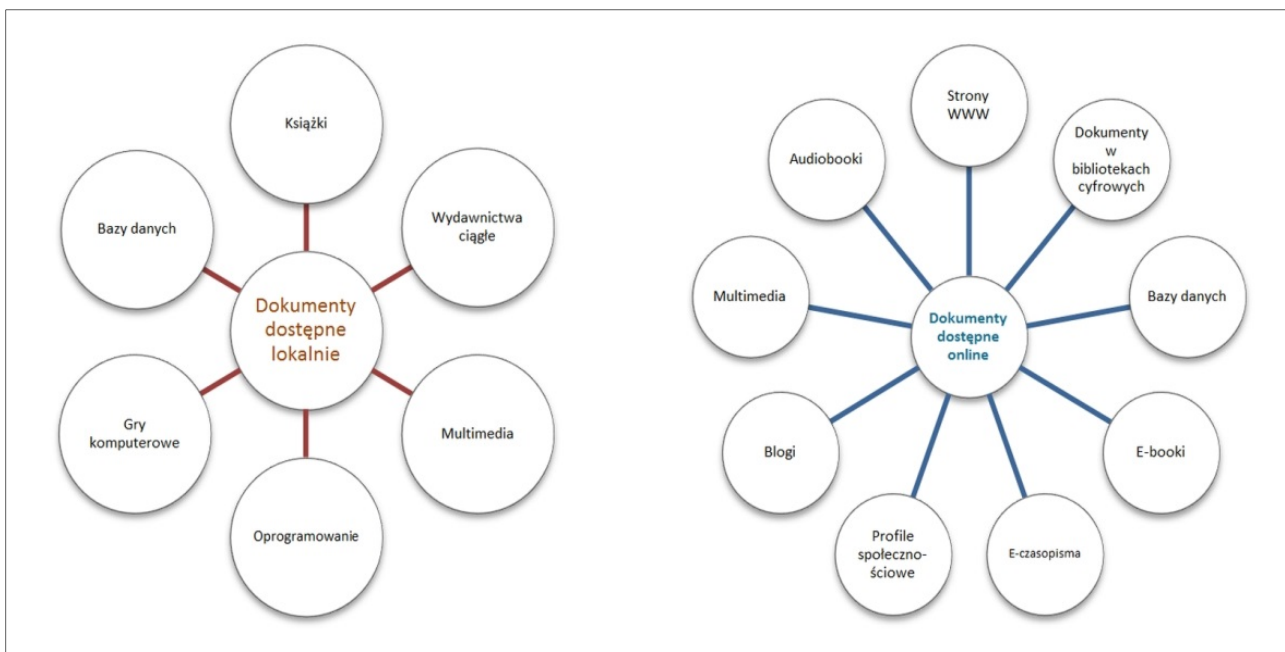
Słowa kluczowe

biblioteki, dane cyfrowe, nośniki danych, formaty plików, przechowywanie danych cyfrowych, archiwizacja Internetu, długoterminowa archiwizacja, projekty archiwizacji danych cyfrowych, Pedagogiczna Biblioteka Cyfrowa, PLATON – Platforma Obsługi Nauki

Biblioteki, obok archiwów, uczelni, instytutów i jednostek naukowych oraz akademickich centrów komputerowych, stają się największymi posiadaczami i jednocześnie wytwórcami danych cyfrowych. Konieczność zabezpieczenia tradycyjnych zbiorów w bibliotekach jest zrozumiała, powszechnie akceptowana i w mniejszym lub większym stopniu realizowana w praktyce. Jednakże świadomość, że cyfrowych nośników nie można przechowywać przez dziesiątki lat w magazynach bibliotecznych, tak jak zasobów drukowanych, nie jest już tak popularna i oczywista. Standardem przechowywania treści przez setki lat był zapis na papierze, natomiast w erze cyfrowej wciąż nie wypracowano uniwersalnych standardów zapisu treści w cyfrowej postaci, które umożliwiłyby ich bezproblemowy odczyt w odległym czasie. Kluczowymi zagrożeniami dla cyfrowych zasobów są: fizyczna nietrwałość nośników danych, nieustanna zmiana formatów zapisu oraz dynamiczny rozwój technologii informatycznych. Dane zapisane dziś na nośniku optycznym lub magnetycznym będą nieczytelne już po kilku czy kilkunastu latach ze względu na fizyczną degradację samych nośników. Z kolei zapis na nawet najtrwalszym nośniku nie zagwarantuje, że dane cyfrowe na nim umieszczone będą możliwe do odczytu przy użyciu sprzętu i oprogramowania dostępnego za kilkadziesiąt lat. Długoterminowe i bezpieczne przechowywanie zgromadzonych zasobów cyfrowych przekracza także możliwości finansowe, organizacyjne i techniczne bibliotek, ponieważ archiwizowanie posiadanych lub wytwarzanych dokumentów cyfrowych powinno odbywać się w sposób bezpieczny, wieczysty, redundantny, skalowalny i geograficznie rozproszony. Zapewnić może go jedynie ogólnokrajowy system powszechnej archiwizacji.

Dane cyfrowe w bibliotekach

Ze względu na sposób udostępniania zasobów cyfrowych w bibliotekach można wyróżnić dokumenty dostępne lokalnie (na FDD, HDD, SSD, CD, DVD, BD, FlashDrive itp.) oraz dokumenty dostępne online [Rys. 1].



Rys. 1. Podział dokumentów cyfrowych w bibliotekach ze względu na sposób udostępniania.

Dokumenty cyfrowe dostępne zarówno lokalnie, jak i te udostępnione online wymagają takich procedur przechowywania, aby ich treść była możliwa do odczytu nie tylko za kilka czy kilkanaście lat, ale najlepiej za kilkaset lat. Cyfrowe zbiory potrzebują więc stałej uwagi i opieki, tak jak zbiory tradycyjne. Tak jak w bibliotekach tradycyjnych, gdzie dbałość o księgozbiór wymaga nakładów finansowych na stworzenie optymalnych warunków długoterminowego przechowywania i dostępu do różnego typu fizycznych obiektów, jak np. książki, rękopisy, mapy, fotografie itp., tak w przypadku obiektów cyfrowych konieczne jest zagwarantowanie odpowiedniego środowiska informatycznego i wypracowanie procedur właściwego przechowywania i dostępu do nich.

Zabezpieczanie danych w bibliotekach najczęściej sprowadza się do wykonywania kopii zapasowych, czyli tzw. *backupu*. Pojęcia tworzenie kopii zapasowych[2] (*backup*) i archiwizacja danych[3] (*archive*), dotyczące zbiorów danych w określonym momencie czasu, są nierzadko traktowane jako synonimy. Często obie te funkcje są łączone ze sobą, więc wydaje się, że mamy do czynienia z tym samym procesem. Różnica między kopią zapasową a archiwum danych zaciera się jeszcze bardziej, gdy kopię zapasową przechowujemy w długim okresie czasu. Podobnie jak kopie zapasowe, tak i zarchiwizowane dane, są rejestrowane i zarządzane w czasie, ale z uwagi na ich niezmiennosc, w dowolnym momencie czasu istnieje tylko jedna ich kopia.

Dane cyfrowe a długoterminowa archiwizacja

Długoterminowa archiwizacja zasobów cyfrowych obejmuje gromadzenie, przechowywanie cyfrowych danych oraz zapewnienie ich użyteczności poprzez dostępność, możliwość odczytu oraz prezentację treści w zrozumiałej dla użytkownika formie, nie tylko aktualnie, ale i w odległej przyszłości. Długoterminowa użyteczność zasobów cyfrowych implikuje systematyczne rozwiązywanie problemów natury technicznej. Najważniejszymi problemami są niska trwałość nośników (utrata pierwotnych właściwości po upływie kilku lub najwyżej kilkudziesięciu lat), starzenie się sprzętu i oprogramowania (starsze wersje wycofywane z rynku i zastępowane nowymi narzędziami, często niekompatybilnymi i niezdolnymi do odtworzenia treści dokumentu w formie czytelnej dla użytkownika) oraz utrata dostępu do dokumentów cyfrowych w wyniku zdarzeń losowych (np. pożar, powódź, awarie). Problemem podobnym do dezaktualizacji formatów danych jest wychodzenie z użycia nośników, takich jak płyty optyczne i magnetoptyczne oraz taśmy i dyski magnetyczne. Jeżeli nawet dane przechowywane na nośniku nie uległy uszkodzeniu i są zapisane w aktualnie dostępnym formacie, to nie posiadamy już urządzeń do odczytu nośnika, który wyszedł z użycia (np. popularne niegdyś dyskietki 5,25"i 3,5"). Oprócz kontroli starzenia się formatów danych zalecane jest również nadzorowanie starzenia się nośników danych.

W przypadku zbiorów cyfrowych nie ukształtowała się jeszcze w bibliotekach powszechna świadomość znaczenia tego problemu. Stąd brak jest wciąż ogólnie przyjętych procedur, które pozwalałyby na opracowanie i wdrożenie strategii długoterminowej archiwizacji danych (*Secondary Storage*) oraz zapewniły finansową i organizacyjną ciągłość dla systemów archiwizacji danych. Bardzo ważnym działaniem jest zabezpieczanie danych poprzez zapisywanie ich w formatach uniwersalnych i otwartych [Tab. 1], szczególnie tych, które zostały stworzone specjalnie na potrzeby długowiecznej archiwizacji np. w wersji formatu PDF – PDF/A[4]. Formatem, który umożliwi nie tylko odczyt, ale również edycję dokumentu jest format ODF[5] w pakiecie LibreOffice[6] o otwartej specyfikacji. Otwarta specyfikacja formatu oznacza, że jego dokumentacja jest dostępna dla wszystkich, a format nie jest zabezpieczony licencjami i patentami. Format otwarty powinien być znormalizowany (np. przez ISO[7]), a treść zapisywana w sposób tekstowy a nie binarny.

FORMATY ZAMKNIĘTE	FORMATY OTWARTE
Opatentowane, chronione licencjami lub utajnione przez producenta formaty plików cyfrowych	Formaty plików cyfrowych o jawnej pełnej dokumentacji struktury formatu, bez ograniczeń licencyjnych w stosowaniu
CDR, PSD, DWG, RAR, WMA, WMV, DOC, XLS, PPT, SWF	HTML, XML, ODF, CSV, PNG, SVG, MKV, FLAC, ePUB

Tab. 1. Przykładowe formaty plików cyfrowych.

Archiwizacja Internetu

Specyficznym rodzajem zasobów cyfrowych są internetowe strony WWW, z których informacje są coraz częściej cytowane i uznawane jako istotne i rzetelne źródła wiedzy.

W dobie powszechnej cyfryzacji społeczeństwo zastępuje papier i druk komputerami, tabletami, smartfonami itp., toteż większość informacji o ważnym znaczeniu historycznym może zostać bezpowrotnie utracona. Witryny internetowe są usuwane lub porzucane, informacje są nietrwałe, często modyfikowane i po pewnym czasie bezpowrotnie znikają. Dynamicznie zmieniające się treści stron internetowych, ich efemeryczny charakter[8], powinny być przesłanką do archiwizacji treści stron WWW. Administratorzy stron WWW zorientowani są w znacznej mierze na utrzymaniu sprawności technicznej bieżącego serwisu i migracji do najnowszych technologii, niż na zachowaniu archiwalnych stron WWW. Zarchiwizowane oficjalne strony internetowe instytucji, w tym bibliotek są bezcennym i wiarygodnym źródłem informacji dla badaczy historii.

Zasoby sieciowe powinny być archiwizowane w sposób przemyślany i zorganizowany. Niezbędnymi działaniami archiwizacyjnymi będzie określenie zasad selekcji (również treściowej), zakresu, czyli „głębokości” kopiowania witryn internetowych, częstotliwości kopiowania stron (codziennie, co tydzień, co miesiąc itp.). Należy również pamiętać o sposobie przyszłego udostępniania zarchiwizowanych stron WWW, co implikuje zachowanie towarzyszącego im środowiska informatycznego.

Światowe projekty archiwizacji

Przykładem instytucji, która podjęła działania archiwizacji Internetu jest British Library[9]. W kwietniu 2013 r. Narodowa Biblioteka Wielkiej Brytanii rozpoczęła archiwizację wszystkich brytyjskich stron internetowych, blogów, newsletterów, e-booków, postów ważnych osób na Twitterze, Facebooku oraz w innych serwisach społecznościowych, aby zachować dla historyków obraz współczesnego społeczeństwa, kultury i informacje o bieżących wydarzeniach. Za cezurę czasową archiwizacji przyjęto rok 2004. Strony internetowe na licencji Creative Commons są archiwizowane bez zgody właściciela, pozostałe natomiast po uzyskaniu zezwolenia posiadaczy witryn. British Library podjęła działania archiwizacyjne na podstawie Legal Deposit Libraries Act[10] z 2003 r., jednakże niezbędne są dalsze regulacje prawne dotyczące gromadzenia historycznych stron WWW. Narzędziem, które wykorzystują użytkownicy UK Web Archive do przeszukiwania treści zarchiwizowanej kolekcji stron WWW jest UK Web Archive N-gram Search[11].

Kolejną inicjatywą wartą przybliżenia jest założona w San Francisco pozarządowa organizacja non profit Internet Archive, która w archiwum pod nazwą WayBack Machine od 1996 r. zgromadziła dotychczas 388 miliardów stron internetowych[12], książki, multimedia, programy komputerowe, czyli głównie obiekty „cyfrowej kultury” – *born digital*. WayBack Machine, podobnie jak tradycyjna biblioteka, oferuje w wolnym dostępie swoje zbiory nie tylko badaczom, nauczycielom akademickim, osobom z dysfunkcją wzroku, ale całemu społeczeństwu. Serwis WayBack Machine umożliwia użytkownikom przeglądanie zmian zachodzących na danej stronie WWW w czasie[13].

Biblioteka Kongresu USA gromadzi również zbiory cyfrowe *born digital*, ale amerykańskie strony internetowe archiwizuje wybiórczo. Gromadzi i zabezpiecza również wpisy na serwisie społecznościowym Twitter[14], jednakże ogromna liczba *tweetów* generuje problemy z ich przechowywaniem i udostępnianiem.

W Bibliotece Watykańskiej do przechowywania i zarządzania zbiorami cyfrowymi wykorzystano technologię *cloud computingu*[15]. Przestrzeń do składowania i zabezpieczania zbiorów cyfrowych Biblioteki Watykańskiej[16] sfinansowano w ramach projektu Information Heritage Initiative[17] dotowanego przez korporację EMC[18]. Cyfrowa Kolekcja Biblioteki Watykańskiej obejmuje 80 000 rękopisów i 8900 inkunabułów.

Projekty archiwizacji w Polsce

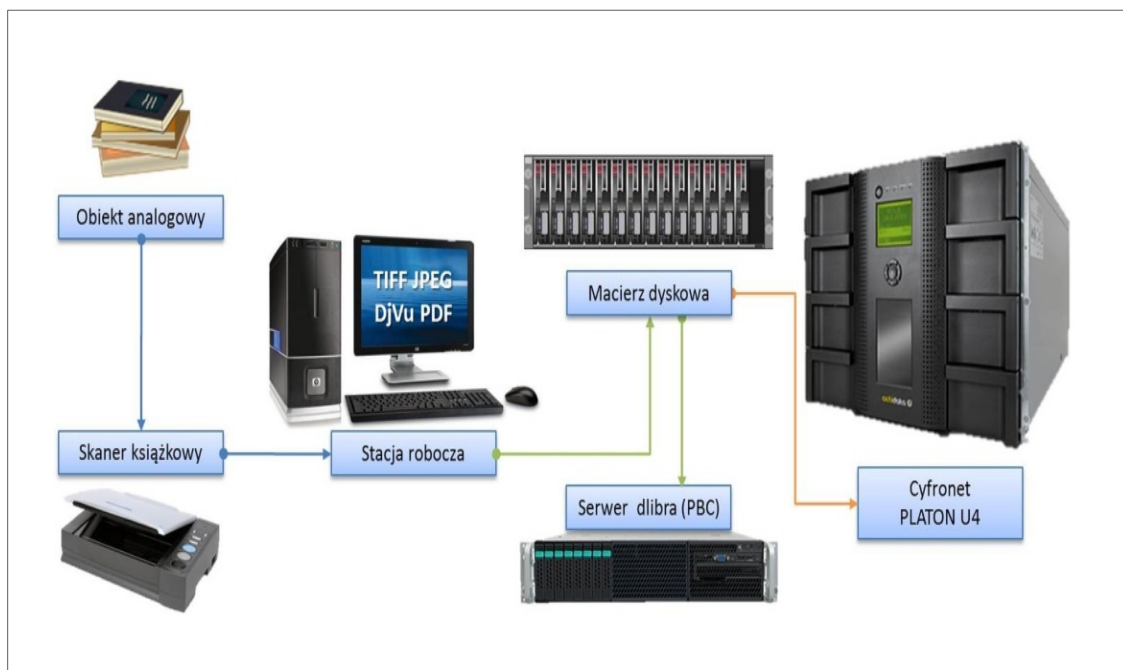
W Polsce cyfrowe informacje o znaczeniu historycznym zabezpiecza Narodowe Archiwum Cyfrowe. Zakres działań Archiwum ogranicza się jednak do archiwizacji stron internetowych urzędów państwowych w domenie rządowej gov.pl. W Archiwum Internetu w Wersji Beta[19] zgromadzono i udostępniono strony najwyższych władz Polski: Prezydenta, Premiera, Sejmu, Senatu, Ministerstwa Kultury i Dziedzictwa Narodowego oraz 34 archiwów i Naczelnej Dyrekcji Archiwów Państwowych, a także Muzeum Narodowego w Krakowie.

Inną ważną inicjatywą jest Repozytorium Dokumentów Elektronicznych Biblioteki Narodowej, uruchomione w kwietniu 2009 r., które gromadzi wyłącznie dokumenty cyfrowe *born digital*, nadsyłane przez wydawców jako egzemplarz obowiązkowy publikacji wydawanych wyłącznie w postaci elektronicznej. Repozytorium zbudowano w oparciu o oprogramowanie DSpace[20] z wydzielonymi dwoma modułami „depozyt” i „publikacje”. „Depozyt” przeznaczono dla wydawców, posługujących się indywidualnymi kontami do samodzielnego przesyłania publikacji szyfrowanymi połączeniami. Moduł „publikacje”, ze względu na ograniczenia wynikające z prawa autorskiego, jest wykorzystywany do udostępniania dokumentów elektronicznych przechowywanych w repozytorium w wewnętrznej sieci komputerowej Biblioteki Narodowej. Każdy obiekt cyfrowy Repozytorium posiada opis bibliograficzny w katalogu komputerowym Biblioteki Narodowej.

Instytucje posiadające biblioteki cyfrowe w Polsce, które nie dysponują znacznymi środkami finansowymi, zmuszone są szukać niskobudżetowych rozwiązań organizacji archiwizacji wytwarzanych zasobów cyfrowych. Archiwizacja efektów pracy zespołu Pedagogicznej Biblioteki Cyfrowej[21] początkowo bazowała na nośnikach optycznych oraz dodatkowo zachowywano pliki master na dyskach lokalnych. Kopiowanie dużych ilości danych na nośniki optyczne było procesem czasochłonnym i niestety zawodnym. Zatem koniecznością stało się każdorazowe weryfikowanie danych po zapisie i ponowne wypalanie błędnie zapisanych dysków, co jeszcze bardziej wydłużało czas archiwizacji i zwiększało koszty. Archiwizacja na nośnikach optycznych oprócz bezdyskusyjnych problemów z trwałością zapisu i późniejszego odczytu, jest czasochłonna i pracochłonna, a stosowana w długim okresie czasu

także kosztowna (rosnąca powierzchnia magazynowa, odświeżanie nośników co kilka lat itp.). Nośniki optyczne zawsze więc traktowano w PBC jako rozwiązanie doraźne i poszukiwano alternatywnej, taniej i bezpiecznej metody archiwizacji materiałów cyfrowych.

W celu usprawnienia działań Zespołu PBC oraz tymczasowego składowania tworzonych danych, wynikających z przypisanych członkom zadań, wprowadzono sieciowy wolumen pośredniczący na dedykowanym serwerze NAS[22], który jednak powinien być zastąpiony macierzą dyskową[23].



Rys. 2. Schemat przepływu danych cyfrowych w Pedagogicznej Bibliotece Cyfrowej.

Natomiast problem długoterminowej archiwizacji dokumentów Pedagogicznej Biblioteki Cyfrowej został rozwiązany dzięki nawiązaniu współpracy z Akademickim Centrum Komputerowym Cyfronet AGH, które jest jednym z uczestników projektu Platformy Obsługi Nauki PLATON[24]. Najważniejszą usługą Projektu z punktu widzenia bibliotek jest Usługa Powszechnej Archiwizacji U4[25], polegająca na udostępnieniu w skali kraju zdalnej archiwizacji i backupu danych cyfrowych.

Możliwość korzystania z profesjonalnie przygotowanej i zarządzanej platformy archiwizacyjnej jest ważnym krokiem do zapewnienia wieczystej archiwizacji dokumentów cyfrowych. Kolejnym działaniem powinno być wdrożenie systemu długoterminowego przechowywania danych źródłowych – dArceo[26], rozwijanego przez Poznańskie Centrum Superkomputerowo Sieciowe w Poznaniu. System dArceo zapewnia zgodność formatów przechowywanych danych z istniejącymi standardami oraz dostarcza narzędzia do ich bezstratnej migracji do nowych formatów.

Konkluzja

W dobie masowej produkcji danych cyfrowych i łatwego dostępu do informacji nie można zapominać o specyficznych wymogach cyfrowych dokumentów w zakresie ich ochrony i przechowywania. Materiał cyfrowy potrzebuje staranniejszego zabezpieczenia niż materiał analogowy, szczególnych warunków przechowywania, gdyż dokumentu cyfrowego nie można umieścić w magazynie bibliotecznym, aby wrócić do niego za kilka, kilkanaście czy kilkadziesiąt lat. Ochrona materiału cyfrowego powinna rozpocząć się już w momencie jego tworzenia i zapisu, a sama konwersja zgromadzonego materiału w postaci cyfrowej jeszcze nie gwarantuje jego użyteczności w długim czasie. Dzięki Projektowi PLATON problem długoterminowej archiwizacji danych cyfrowych został częściowo rozwiązany dla środowiska akademickiego. Natomiast wciąż brak jest długofalowej strategii zabezpieczenia i archiwizacji materiałów cyfrowych wydawanych oficjalnie nie tylko na nośnikach fizycznych, ale również tych publikowanych w Internecie. Aktualnie publiczne środki chętnie przekazywane są na realizację małych i dużych projektów digitalizacyjnych bibliotek, natomiast sama kwestia długoterminowego, fachowego zabezpieczenia już wytworzonych danych cyfrowych nadal nie znajduje należytego finansowego wsparcia.

Przypisy:

- [1] Su-Shing Chen, *The paradox of digital preservation*, "Computer", 2001, Vol. 34, nr 3, s. 24-28, doi:10.1109/2.910890.
- [2] Kopie zapasowe są definiowane jako rzuty zbioru danych wykonanych w określonym momencie czasu, zapisane w powszechnie dostępnym formacie i monitorowane przez cały czas wykorzystywania w systemie tworzącym kopie zapasowe.
- [3] Archiwizacja danych polega na przeniesieniu zbioru oryginalnych stałych danych wg określonych kryteriów z jednej lokalizacji do innej, najlepiej oddalonej geograficznie, w celu przechowywania na nośnikach odpornych na degradację w długim okresie czasu.
- [4] PDF/A (archiwalny PDF) to format pliku służący do długoterminowej archiwizacji danych elektronicznych, przyjęty jako standard przez Międzynarodową Organizację Normalizacyjną, nr ISO 19005. PDF/A zawiera oprócz samego tekstu również informację o użytych w dokumencie czcionkach, grafikach wektorowych i rastrowych, przypisach, komentarzach itp.
- [5] ODF (Open Document Format) jest otwartym standardem ISO formatu plików pakietów biurowych.
- [6] LibreOffice jest wielofunkcyjnym *open source* pakietem oprogramowania biurowego, powstałym na bazie kodu OpenOffice.org.
- [7] ISO (International Organization for Standardization) jest pozarządową Międzynarodową Organizacją Normalizacyjną, respektowanie norm ISO jest dobrowolne. Autorytet ISO wynika z międzynarodowej reprezentacji.
- [8] Średni czas życia strony WWW szacowany jest na 75 dni.
- [9] *UK Web Archive* [online], 2014 [dostęp: 2014-03-10]. Dostępny w World Wide Web: <http://www.webarchive.org.uk/>.
- [10] *Legal Deposit Libraries Act* z 2003 r. jest ustawą Parlamentu brytyjskiego uchwaloną w celu aktualizacji ustawodawstwa w zakresie egzemplarza obowiązkowego w erze dokumentów cyfrowych.
- [11] UK Web Archive N-gram Search pozwala przeszukiwać treść wszystkich zgromadzonych zasobów i wizualizować częstotliwość występowania słów kluczowych na osi czasu. UK Web Archive udostępnia też możliwość pełnotekstowego przeszukiwania bazy zgromadzonych zasobów.
- [12] *Internet Archive* [online], 2014 [dostęp: 2014-03-10]. Dostępny w World Wide Web: <https://archive.org/>.
- [13] Funkcjonalność tą Internet Archive nazwał *three dimensional index*.

[14] Jednym z ważnych zarchiwizowanych tweetów jest wpis Baracka Obamy z 2008 r. po wygranych wyborach prezydenckich: „Właśnie stworzyliśmy historię. Dziękuję”.

[15] *Cloud computing* (chmura obliczeniowa) jest modelem przetwarzania danych opartym na wynajmie usług dostarczonych przez usługodawcę w zakresie: infrastruktury technicznej (np. komputery, pamięci, sprzęt sieciowy, sieci), platform programowych (np. bazy danych, systemów identyfikacji, narzędzi do budowy SI, modułów uruchomieniowych), aplikacji użytkowych (np. systemy finansowe, kadrowe, statystyczne, programy graficzne).

[16] Przestrzeń internetowa dla zbiorów Biblioteki Watykańskiej wynosi 2,8 petabajta (1 PB = 1000 TB). Zbiory cyfrowe Biblioteki Watykańskiej są dostępne pod adresem: *Biblioteca Apostolica Vaticana* [online], 2014 [dostęp: 2014-03-10]. Dostępny w World Wide Web: http://www.vaticanlibrary.va/home.php?pag=mss_digitalizzati&BC=11.

[17] Information Heritage Initiative jest filantropijnym projektem firmy EMC, którego celem jest uczynienie z historycznych dokumentów i artefaktów kulturowych dobra wspólnego, dostępnego dla badań naukowych i edukacji przez Internet. Strona projektu: *The Information Heritage Initiative* [online], 2014 [dostęp: 2014-03-10]. Dostępny w World Wide Web: http://www.emc.com/leadership/features/information_heritage_initiative.htm.

[18] EMC jest amerykańską korporacją dostarczającą produkty, usługi i rozwiązania z dziedziny zarządzania informacją od przechowywania danych, archiwizacji, kopii bezpieczeństwa, deduplikacji, po kompleksowe rozwiązania bezpieczeństwa i zarządzania treścią. Obecna nazwa korporacji nawiązuje do równania Einsteina: $E=mc^2$.

[19] *Archiwum Internetu Beta* [online], 2014 [dostęp: 2014-03-10]. Dostępny w World Wide Web: <http://www.archiwuminternetu.nac.gov.pl/>.

[20] DSpace jest oprogramowaniem *open source* zaprojektowanym przez Massachusetts Institute of Technology we współpracy z Hewlett Packard Company. Obecnie jest najbardziej popularnym narzędziem do tworzenia repozytoriów, korzysta z niego ok. 380 instytucji na całym świecie.

[21] Autorzy artykułu są koordynatorami i administratorami Pedagogicznej Biblioteki Cyfrowej.

[22] NAS (Network Attached Storage) to technologia umożliwiająca skonfigurowanie dostępu do danych zgromadzonych w jednym miejscu z różnych punktów sieci komputerowej, dzięki podłączeniu zasobów pamięci dyskowych do sieci komputerowej.

[23] Macierz dyskowa to wyspecjalizowane urządzenie pamięci masowej, którego głównymi elementami są twarde dyski (od kilku do kilkuset) zgrupowane w układach RAID (Redundant Array of Independent Disks).

[24] Celem projektu Platforma Obsługi Nauki PLATON jest wdrożenie pięciu nowoczesnych usług działających w oparciu o sieć Pionier: videokonferencji, eduroam, kampusowych, powszechnej archiwizacji, naukowej interaktywnej telewizji, dostępnych dla środowiska naukowego w Polsce. Nazwa projektu PLATON nawiązuje do greckiego filozofa, a sam projekt wskazuje na sieci komputerowe jako medium niosące nowe możliwości i usługi.

Strona projektu: *PLATON* [online], 2014 [dostęp: 2014-03-10]. Dostępny w World Wide Web: <http://www.platon.pionier.net.pl/online>.

[25] Usługa Powszechnej Archiwizacji Danych (U4) realizowana jest w ramach Projektu PLATON przez konsorcjum specjalistycznych instytucji posiadających infrastrukturę sprzętową, oprogramowanie oraz wykwalifikowany personel. Węzły przechowywania danych oraz węzły usługowe są rozproszone geograficznie w różnych miastach Polski i połączone z optycznymi sieciami miejskimi oraz siecią krajową. Dane są archiwizowane z nieograniczonym w czasie dostępem online do archiwum. Usługa U4 gwarantuje bezpieczeństwo funkcjonowania jednostek akademickich w całym kraju, przechowując generowane dokumenty cyfrowe np. wyniki eksperymentów naukowych, badań, materiałów edukacyjnych, dokumentów bibliotek cyfrowych i in.

[26] dArceo jest systemem długoterminowego przechowywania danych źródłowych (np. plików master) dla plików tekstowych, graficznych i audiowizualnych. System umożliwia migrację danych źródłowych zgodnie z założeniami modelu OAIS (Open Archival Information System). dArceo oferuje funkcje: przechowywania i wersjonowania danych, zarządzania metadanymi, repozytorium OAI-PMH, przekształcania danych, monitorowania danych, wspólną przestrzeń przekształcania danych. dArceo może współpracować z usługą Powszechnej Archiwizacji U4 Projektu PLATON. Strona systemu: *dArceo* [online], 2014 [dostęp: 2014-03-10]. Dostępny w World Wide Web: <https://dingo.psnc.pl/darceo>.

Bibliografia:

- [1] Brzeźniak M., *Usługa Powszechnej Archiwizacji i jej zastosowanie w bibliotekach naukowych dla zabezpieczenia i archiwizacji danych*, „EBIB”, [online], 2010, nr 6 [dostęp: 2014-03-10]. Dostępny w World Wide Web: <http://www.ebib.pl/2010/115/a.php?brzezniak>.
- [2] Januszko-Szakiel A., *Długoterminowa archiwizacja zasobów cyfrowych dla polskich bibliotek (program)*, „Przegląd Biblioteczny”, 2011, z. 2, s. 21-46.
- [3] Januszko-Szakiel A., *Długoterminowa archiwizacja zasobów cyfrowych w świadomości pracowników polskich bibliotek* [online], 2009 [dostęp: 2014-03-10]. Dostępny w World Wide Web: <http://lib.psnc.pl/Content/370/11-Januszko-ER.pdf>.

- [4] Januszko-Szakiel A., *Nowy wymiar zabezpieczania elektronicznych zasobów*, „EBIB”, [online], 2010, nr 6 [dostęp: 2014-03-10]. Dostępny w World Wide Web: <http://www.ebib.pl/2010/115/a.php?januszko>.
- [5] Januszko-Szakiel A., *Wiarygodność archiwów cyfrowych*, „Przegląd Biblioteczny”, 2009, z. 3, s. 325-347.
- [6] Lorie R. A., *Long-term archiving of digital information* [online], 2000 [dostęp: 2014-03-10]. Dostępny w World Wide Web: [http://domino.watson.ibm.com/library/CyberDig.nsf/papers/BE2A2B188544DF2C8525690D00517082/\\$File/RJ10185.pdf](http://domino.watson.ibm.com/library/CyberDig.nsf/papers/BE2A2B188544DF2C8525690D00517082/$File/RJ10185.pdf).
- [7] Nelson S., *Profesjonalne tworzenie kopii zapasowych i odzyskiwanie danych*, Gliwice 2012.
- [8] *PIONIER Magazine : Polish Optical Internet* [online], 2011, nr 1 [dostęp: 2014-03-10]. Dostępny w World Wide Web: http://www.pionier.net.pl/magazine/pl/artykuly/1365/PLATON_specjalnie.html.
- [9] *Platforma Obsługi Nauki PLATON* [online], 2014 [dostęp: 2014-03-10]. Dostępny w World Wide Web: <http://www.platon.pionier.net.pl/online>.
- [10] Witczak D., Sobkowiak K., Panaś K., *W oczekiwaniu na PLATONa : współpraca Pedagogicznej Biblioteki Cyfrowej z Akademickim Centrum Komputerowym Cyfronet AGH w zakresie długoterminowej archiwizacji danych*. In Januszko-Szakiel A. (red.), *Wokół zagadnień trwałej ochrony zasobów cyfrowych*, Kraków, 2013, s. 89-99.

Informacja o autorach:

mgr Dorota Witczak – kierownik Oddziału Komputeryzacji i Digitalizacji Zbiorów Biblioteki Głównej Uniwersytetu Pedagogicznego im. KEN w Krakowie, e-mail: dorota.witczak@libpost.up.krakow.pl, tel. 12 662 63 74.

Krzysztof Sobkowiak – pracownik Oddziału Komputeryzacji i Digitalizacji Zbiorów Biblioteki Głównej Uniwersytetu Pedagogicznego im. KEN w Krakowie, e-mail: krzysztof@libpost.up.krakow.pl, tel. 12 662 63 74.